

Transitions between spatial attractors in place-cell networks

R. Monasson, S. Rosay

Laboratoire de Physique Théorique de l'ENS, CNRS & UPMC, 24 rue Lhomond, 75005 Paris, France

(Dated: July 22, 2015)

The spontaneous transitions between D -dimensional spatial maps in an attractor neural network are studied. Two scenarios for the transition from one map to another are found, depending on the level of noise: (1) through a mixed state, partly localized in both maps around positions where the maps are most similar; (2) through a weakly localized state in one of the two maps, followed by a condensation in the arrival map. Our predictions are confirmed by numerical simulations, and qualitatively compared to recent recordings of hippocampal place cells during quick-environment-changing experiments in rats.

In 1982 J.J. Hopfield proposed a neural network model for auto-associative memories, in which specific configurations (patterns) of the activity could be stored through an adequate choice of the interactions between the neurons, modeled as binary units $\sigma = 0, 1$ [1]. Starting from an initial configuration partially resembling one pattern the network configuration dynamically evolves until a fixed point, coinciding almost exactly with the stored pattern, is reached [2]. In this paper we consider an extension of the Hopfield model, in which the attractors stored are continuous and D -dimensional (with $D \geq 1$), rather than discrete fixed points ($D = 0$), and discuss the existence and the mechanisms of spontaneous transitions between those attractors. Besides its intrinsic interest from a statistical mechanics point of view our study is motivated by the observation of abrupt transitions between the representations of space in the brain [3], in particular in quick-environment-changing experiments on rats [4].

Continuous attractors are not unusual in statistical physics. An illustration is given by the Lebowitz-Penrose theory of the liquid-vapor transition [5]. Consider a D -dimensional lattice, whose N sites \vec{x}_i can be occupied by a particle ($\sigma_i = 1$), or left empty ($\sigma_i = 0$). The energy of a configuration $\{\sigma_i\}$ is given by the Ising-like Hamiltonian

$$E[\{\sigma_i\}, \{\vec{x}_i\}] = - \sum_{i < j} J(|\vec{x}_i - \vec{x}_j|) \sigma_i \sigma_j, \quad (1)$$

where J is a positive and decaying function of its argument, *i.e.* of the distance between sites. At fixed number of particles and low enough temperature translation invariance on the lattice is spontaneously broken: particles tend to cluster in the \vec{x} -space, and form a high density region (liquid drop) surrounded by a low-density vapor. The density profile of this ‘bump’ of particles hardly fluctuates, but its position can freely move on the lattice, and defines a collective coordinate for the microscopic configuration of particles.

From the neuroscience point of view, the existence of a collective coordinate, weakly sensitive to the high stochasticity of the microscopic units, is central to population coding theory [6]. Following the seminal discovery

of ‘place cells’ in a brain area called hippocampus [7], continuous attractors have been proposed as the basic principle of the coding for position in space [8]. The model we consider here goes as follows [9]. In a rat moving freely in a given environment, a place cell i becomes active ($\sigma_i = 1$) when the rat is at a specific location in the environment, called place field and centered in \vec{x}_i , and silent otherwise ($\sigma_i = 0$). Place cells i, j with overlapping place fields, *i.e.* such that the distance between \vec{x}_i and \vec{x}_j is small enough, may be simultaneously active, and have a tendency to strengthen their connection during the exploration of the environment. The potentiation of the couplings between coactive neurons is called Hebbian learning, an important paradigm in auto-associative memories. In addition to the local excitation, global inhibition across the network keeps the fraction of active units constant. As a result the probability of a place-cell activity configuration $\{\sigma_i\}$ would formally coincide with the Gibbs measure associated to model (1), with a temperature T dependent on the neural noise.

When the rat explores a new environment hippocampal place cells undergo a process called remapping, in which place field locations are randomly reallocated [10]. A simple model of remapping consists in randomly permuting the indices of the place field locations, *i.e.* the center of the place field attached to place cell i becomes $\vec{x}_{\pi(i)}$, where π is a random permutation defining the spatial map of the environment. We assume that the contributions to the interactions of the different maps add up, and obtain the Hopfield-like hamiltonian

$$E[\{\sigma_i\}, \{\vec{x}_i\}, \{\pi^\ell\}] = - \sum_{i < j} J_{ij} \sigma_i \sigma_j, \quad (2)$$

where the couplings are

$$J_{ij} = \sum_{\ell=1}^L J(|\vec{x}_{\pi^\ell(i)} - \vec{x}_{\pi^\ell(j)}|), \quad (3)$$

L is the number of environments, and π^ℓ the permutation in the ℓ^{th} environment ($\ell = 1, \dots, L$). This J_{ij} -matrix exhibits a ‘small-world’-like topology [11]: the couplings due to an environment, say, $\ell = \ell_0$ in (3), connect neurons

with place fields close to each other in this environment, while the other environments ($\ell \neq \ell_0$) contribute long-range and random connections. Despite this structural disorder, in model (2) and in similar rate-based models with continuous units [12–15], the activity can be localized in any of the environments, *i.e.* the active neurons have neighboring place fields in one map, provided the load $\alpha = L/N$ and the temperature T are not too large.

The purpose of this paper is to study the transitions between different maps, *i.e.* how the population activity can abruptly jump from being localized in one map to another one. For definiteness we choose the N place-fields centers to be the nodes \vec{x}_i of a regular D -dimensional cubic lattice. The coupling function in (1) is set to $J(r) = \frac{1}{N}$ if $r < r_c$, and 0 otherwise, where r measures the distance on the grid; the cut-off distance r_c is such that each neuron is connected to its wN closest neighbors in each map. The fraction of active neurons is fixed to the value f . We report in Fig. 1 the outcome of Monte Carlo simulations of model (2) with $L = 2$ maps, referred to as A and B, in dimension $D = 1$ (See Supplemental Material at [url] for a simulation in $D = 2$ dimensions). The bump of activity diffuses within one map with little deformation, and sporadically jumps from map A to B, and back. We study below the mechanisms underlying those transitions, remained poorly understood so far.

To capture the typical properties of model (2) we compute its free energy under the constraint that the average activities of neurons whose place fields are centered in \vec{x} in map A and in \vec{y} in map B are equal to, respectively, $\rho^A(\vec{x})$ and $\rho^B(\vec{y})$. We use the replica method to average the free energy over the random permutations π^ℓ . The outcome, within the replica symmetric hypothesis and for $N \rightarrow \infty$, is the free energy per neuron:

$$\begin{aligned} \mathcal{F} = & -\frac{1}{2} \int d\vec{x} \int d\vec{x}' \rho^A(\vec{x}) J(|\vec{x} - \vec{x}'|) \rho^A(\vec{x}') \\ & - \frac{1}{2} \int d\vec{y} \int d\vec{y}' \rho^B(\vec{y}) J(|\vec{y} - \vec{y}'|) \rho^B(\vec{y}') + \frac{\alpha\beta}{2} r(f - q) \\ & + \int d\vec{x} \mu^A(\vec{x}) \rho^A(\vec{x}) + \int d\vec{y} \mu^B(\vec{y}) \rho^B(\vec{y}) - \eta f \\ & - \alpha \sum_{\vec{k} \neq \vec{0}} \left[\frac{(q - f^2) \Lambda_{\vec{k}}}{1 - \beta(f - q) \Lambda_{\vec{k}}} - T \log(1 - \beta(f - q) \Lambda_{\vec{k}}) \right] \\ & - T \int d\vec{x} d\vec{y} D z \log(1 + e^{\beta(\mu^A(\vec{x}) + \mu^B(\vec{y}) + z\sqrt{\alpha r} - \eta)}) , \end{aligned} \quad (4)$$

where $\beta = 1/T$, $Dz \equiv \exp(-z^2/2)/\sqrt{2\pi}$, $\Lambda_{\vec{k}} \equiv \prod_{\mu=1}^D \sin(k_\mu \pi w^{1/D})/(\pi k_\mu)$ are the eigenvalues of the J matrix, and the components k_μ are positive integer numbers. Fields $\mu^A(\vec{x})$, $\mu^B(\vec{y})$, and parameter r are conjugated to, respectively, the densities $\rho^A(\vec{x})$, $\rho^B(\vec{y})$, and the Edwards-Anderson overlap $q = \frac{1}{N} \sum_i \langle \sigma_i \rangle^2$, where $\langle \cdot \rangle$ denotes the Gibbs average with energy (2) at temperature T , and $\overline{(\cdot)}$ is the average over the random permutations. Parameter η is chosen to enforce the fixed-activity con-

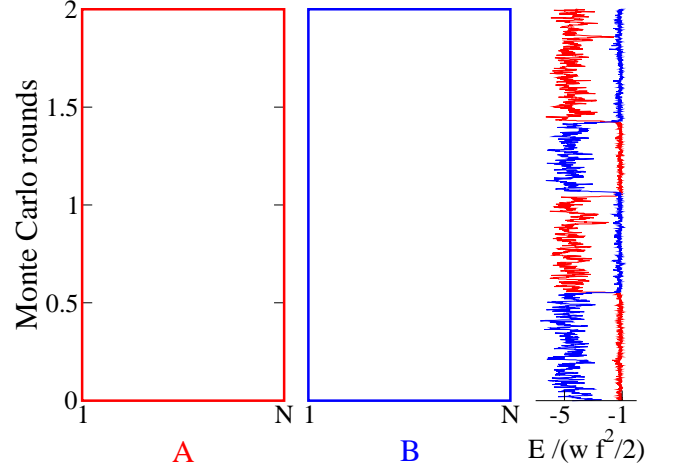


FIG. 1: (Color online) Time evolution of a network with $N = 1000$ neurons, and $L = 2$ one-dimensional maps. Each black dot represents an active neuron σ_i . Panels A and B show the same data, up to a permutation of the neuron indices i to sort place field centers in increasing order in the corresponding map (x -axis). Right: contributions to the total energy due to each map, divided by $\frac{1}{2} f^2 w$ (absolute value of the PM energy). Parameter values: $T = 0.007$, $f = 0.1$, $w = 0.05$. Each MC round includes 10^6 steps.

straint: $\int d\vec{x} \rho^A(\vec{x}) = \int d\vec{y} \rho^B(\vec{y}) = f$.

Minimization of \mathcal{F} allows us to determine the properties of the ‘clump’ phase (CL), in which the activity has a bump-like profile in one map, and is flat ($= f$) in the other one, see Fig. 2(a). Details about the activity profile for a given realization of the maps are shown in Supplemental Material. This CL phase corresponds to the retrieval of one map. The phase diagram in the α, T plane is shown in Fig. 3(a), and also includes the paramagnetic (PM) phase at high T and the spin-glass (SG) phase at high α , in which no map can be retrieved [9]. In the PM phase, the neural noise is large enough to wipe out any interaction effect: the average activity $\langle \sigma_i \rangle$ of

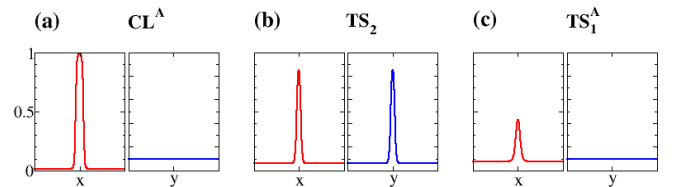


FIG. 2: (Color online) Activity profiles $\rho^A(x)$ (red, left panels) and $\rho^B(y)$ (blue, right panels) computed from replica theory for (a) the phase CL^A ($T = 0.006$), (b) the transition state TS_2 ($T = 0.006$), (c) the transition state TS_1^A ($T = 0.007$). Profiles attached to CL^A and TS_1^A are obtained from CL^A and TS_1^A by swapping ρ^A and ρ^B . Densities were discretized over 201 bins. Parameter values: $f = 0.1$, $w = 0.05$, $L = 2$ one-dimensional maps.

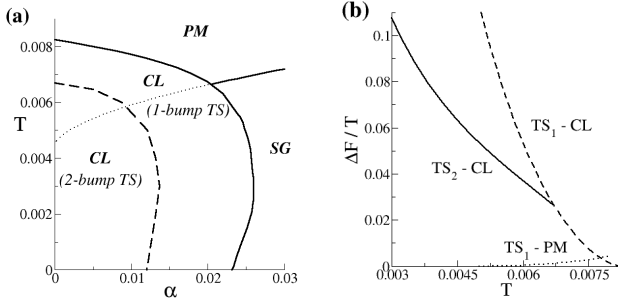


FIG. 3: **(a)** Region of stability of the CL phase and PM-SG boundary in the (α, T) plane (thick lines); the PM and SG phases coexist with the CL phase, respectively, above and below the dotted line. Dashed line: boundary between 1- and 2-bump transition state (TS) scenarios. **(b)** Theoretical barriers $\Delta\mathcal{F}/T$ vs. temperature T for $L = 2$ environments. Same parameter values as in Fig. 2.

each neuron coincides with the global average activity, f . In the SG phase, activities are non uniform ($\langle\sigma_i\rangle \neq f$) for a given realization of the maps, but reflect the random crosstalk between the maps: they do not code for a well-defined position in any environment.

To understand the transition mechanisms we look for saddle-points of \mathcal{F} , through which the transition pathway connecting CL^A to CL^B will pass with minimal free-energy cost $N \Delta\mathcal{F}$. According to nucleation theory [16] we expect those transition states (TS) to be unstable along the transition pathway, and stable along the other directions. We identify two types of TS, referred to as 1 and 2 respectively, depending on the number of maps in which the activity is localized at the saddle-point (Fig. 2(b)&(c)). The corresponding transition pathways are schematized in Fig. 4. Which type of TS is chosen by the system depends on the values of α and T , see Fig. 3(a).

At low enough temperature the transition pathway passes through a transition state, TS_2 , where the activity is equally localized in both maps (Fig. 2(b)). TS_2 is a minimum of the free-energy \mathcal{F} in the symmetric subspace $\rho^A = \rho^B$, and is unstable against one transverse fluctuation mode (Fig. 4(a)). The barrier $\Delta\mathcal{F}$ is the difference between the free-energies of TS_2 and CL [16], and is shown as a function of temperature in the case $L = 2$ in Fig. 3(b). For a given realization of the maps, the lowest free-energy barrier will be achieved by centering the bumps around positions \vec{x} in A and \vec{y} in B, such that the maps are locally similar, *i.e.* such that the adjacency matrices of the maps locally coincide. We define the local resemblance between maps A and B at respective positions \vec{x} and \vec{y} through

$$\text{Res}_{AB}(\vec{x}, \vec{y}) \equiv \frac{1}{N} \sum_{i=1}^N \rho(\vec{x} - \vec{x}_{\pi^A(i)}) \rho(\vec{y} - \vec{x}_{\pi^B(i)}) \quad , \quad (5)$$

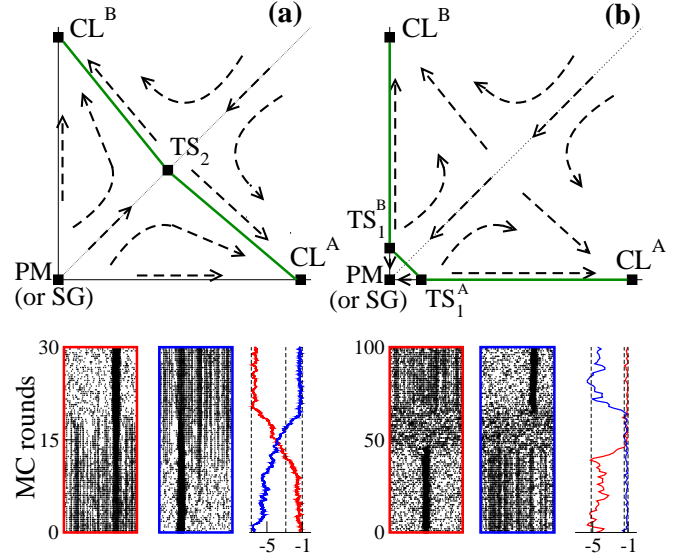


FIG. 4: (Color online) Top: representative points of the CL^A , CL^B , PM or SG phases, and of the saddle-points TS_2 (left), TS_1^A and TS_1^B (right). Coordinates correspond to the averaged squared densities in maps A (x -axis) and B (y -axis), and arrows represent the stability of the phases and saddle-points. The transition pathways are sketched by the green straight lines. Bottom: Monte Carlo simulations illustrating the transition scenarios, and energy contributions due to each map in units of $f^2 w/2$ (x -axis as in Fig. 1); Dotted lines, from left to right: theoretical values of the energies in the CL, TS, PM phases. Parameter values: $N = 1000$, $f = 0.1$, $w = 0.05$, $L = 2$, $T = 0.006$ (left) and $T = 0.007$ (right).

where ρ denotes the bump profile of TS_2 in Fig. 2, common to the two maps. $\text{Res}_{AB}(x, y)$ is shown in Fig. 5(a) for two randomly drawn one-dimensional maps, and compared in Fig. 5(b) to the number of transitions, starting at position x in A and ending at position y in B (or vice-versa), observed in Monte Carlo simulations. Both quantities are strongly correlated, showing that transitions take place at positions where maps are most similar, as intuited in [17]. In addition, those ‘gateway’ positions are energetically favorable, and kinetically trap the activity bump (Fig. 1), hence making transitions more likely to happen.

At high temperatures or loads no saddle-point of \mathcal{F} localized in two maps can be found. At such temperatures or loads the PM or the SG phase coexists with CL (Fig. 3(a)). The PM/SG phase is separated from CL^A by a transition state TS_1^A (Figs. 2(c) & 4(b)), and from CL^B by a transition state TS_1^B . The transition pathway goes from CL^A to TS_1^A , then to TS_1^B at constant \mathcal{F} level, before reaching CL^B (Fig. 4(b)). The barrier $\Delta\mathcal{F}$ is given by the difference between the free energies of CL and TS_1 , and is shown in Fig. 3(b). Note that at intermediate T or α values PM or SG are very shallow local minima of \mathcal{F} (Fig. 3(b), dotted line). The system is likely to transiently visit PM or SG from TS_1^A or TS_1^B . The activity

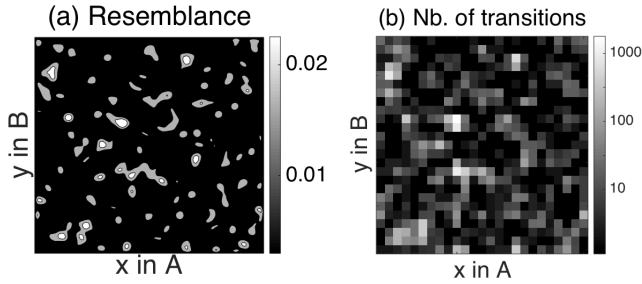


FIG. 5: (a) Local resemblance of maps (contour lines of $\text{Res}_{AB}(x, y)$), and (b) number of transitions between positions x in A and y in B observed in Monte Carlo simulation, for two randomly drawn one-dimensional maps A and B. Parameters: $T = 0.006$, $f = 0.1$, $w = 0.05$, $L = 2$, $N = 333$. Transitions were counted over 10^6 rounds of $10N$ MC steps, with starting/ending positions binned in 25 cells in each map.

configurations are then delocalized in both maps, before eventually condensating into the CL phase. Model (2) defines a dilute ferromagnet with inhomogeneities in the interaction network, see (3). Bumps of neural activity are likely to melt, and TS_1 -based transitions to take place, where the network is less dense [18].

Two-clump (TS_2) and one-clump/delocalized (TS_1) transition scenarios are observed in simulations as reported in Fig. 4 and Supplemental Material, Figs. 4, 5&6, whether the CL phase coexist with the PM (small loads α) or the SG (moderate loads) phases, see Fig. 3(a). The two scenarios also coexist over a range of temperatures (Fig. 3(b)), and may be alternatively observed in finite-size simulations. As an illustration the second transition in Fig. 1 is of type 1 (at $\simeq 1.05 \cdot 10^6$ MC steps), while all three other transitions are of type 2. The boundary line along which both scenarios have equal free-energy barriers is shown in Fig. 3(a). Simulations confirm that the transition rate increases with α and T , and decreases exponentially with N (Supplemental Material, Figs. 1&2).

It is interesting to compare the scenarios above to the experiment by K. Jezek *et al.* [4], in which a rat was trained to learn two environments, A and B, differing by their light conditions. The activity of $\simeq 30$ recorded place cells was observed to rapidly change from being typical of environment A to being typical of environment B, or vice-versa, either spontaneously, or as a result of a light switch. During light cue-induced transitions mixed states, correlated with the representative activities of both maps were observed for a few seconds (Figs. 3a&b, and Supplementary Figs. 6&7b in [4]). Spontaneous transitions were also found to take place, in correspondence to mixed states, or to neural configurations seemingly uncorrelated with A and B, see Fig. 3c in [4]. Those findings are qualitatively compatible with our two transition scenarios. A quantitative comparison of our model with the neural activity in the CA3 and CA1 areas recorded in [4] will be reported in a forthcoming publication.

Our work could be extended along several lines, e.g. to study the consequences of rhythms, such as the $\simeq 8$ Hz Theta oscillations, believed to be very important for the exploration of the space of neural configurations [20], and, hence, for transitions [4, 21]. In addition, we have assumed here, for the sake of mathematical tractability, that the coupling matrix in each map was homogeneous (the number of neighbors of each neuron is uniform across the population), a result of perfect exploration and learning of the environment. In reality, imperfect learning, irregularities in the positions and shapes of place fields, and the sparse activity of place cells in CA1 [22], and even more so in CA3 will concur to produce heterogeneities in the coupling matrix. Numerical simulations suggest, however, that the mechanisms of transitions we have analytically unveiled in the homogeneous case are unaltered in the presence of heterogeneities (Supplemental Material). Last of all, the notion of space itself could be revisited. The ‘overdispersion’ of the activity of place cells [23], its dependence on task and context [24], ... suggest that place cells code for ‘positions’ in a very high-dimensional space, whose projections onto the physical space are the commonly defined place fields. Extending our study to the case of generic metric spaces could be very interesting, and shed light on the existence of fast transitions between task-related maps [25] and, more generally, on the attractor hypothesis as a principle governing the activity of the brain.

Acknowledgements. We are grateful to J.J. Hopfield and K. Jezek for very useful discussions. R.M. acknowledges financial support from the [EU]-FP7 FET OPEN project Enlightenment 284801 and the CNRS-InphyNiTi project INFERNEUR.

-
- [1] J.J. Hopfield, *Proc. Nat. Acad. Sci. USA* **79**, 2554 (1982).
 - [2] D.J. Amit, H. Gutfreund, H. Sompolinsky, *Phys. Rev. Lett.* **55**, 1530 (1985).
 - [3] T.J. Wills *et al.*, *Science* **308**, 873 (2005).
 - [4] K. Jezek *et al.*, *Nature* **478**, 246 (2011).
 - [5] J. Lebowitz, O. Penrose, *J. Math. Phys.* **7**, 98 (1966).
 - [6] S.-I. Amari, *Bio. Cyber.* **27**, 77 (1977).
 - [7] J. O’Keefe, J. Dostrovsky, *Brain Res.* **34**, 171 (1971).
 - [8] M.V. Tsodyks, T.J. Sejnowski, *International J. Neural Syst.* **6**, 81 (1995).
 - [9] R. Monasson, S. Rosay, *Phys. Rev. E* **87**, 062813 (2013).
 - [10] J.L. Kubie, R.U. Muller, *Hippocampus* **1**, 240 (1991).
 - [11] J.D. Watts, S.H. Strogatz, *Nature* **393**, 440 (1998).
 - [12] A. Samsonovich, B.L. McNaughton, *J. Neurosci.* **17**, 5900 (1997).
 - [13] F.P. Battaglia, A. Treves, *Phys. Rev. E* **58**, 7738 (1998).
 - [14] M.V. Tsodyks, *Hippocampus* **9**, 481 (1999).
 - [15] J.J. Hopfield, *Proc. Nat. Acad. Sci. USA* **107**, 1648 (2010).
 - [16] J.S. Langer, *Annals of Physics* **54**, 258 (1969).
 - [17] B.L. McNaughton *et al.*, *J. Exp. Biology* **199**, 173 (1996).
 - [18] R.B. Griffiths, *Phys. Rev. Lett.* **23**, 17 (1969).

- [19] R. Monasson, S. Rosay, *Phys. Rev. E* **89**, 032803 (2014).
- [20] G. Buzsaki, *Rhythms of the brain*, Oxford University Press (2011).
- [21] F. Stella, A. Treves, *Neural Plasticity*, 683961 (2011).
- [22] L.T. Thompson, P.J. Best, *J. Neurosci.* **9**, 2382 (1989).
- [23] A.A. Fenton, R.U. Muller, *Proc. Natl. Acad. Sci. USA* **95**, 3182 (1998).
- [24] D.M. Smith, S.J.Y. Mizumori, *Hippocampus* **16**, 716 (2006).
- [25] J. Jackson, A.D. Redish, *Hippocampus* **17**, 1209 (2007).
- [26] See Supplemental Material [url].